STRUCTURE-BASED DRUG DISCOVERY AGAINST THE CASEINOLYTIC PROTEASE FROM *MYCOBACTERIUM TUBERCULOSIS*

An Undergraduate Research Scholars Thesis

by

EMILY GRACE TRANCHINA

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisor: Dr. James C. Sacchettini

May 2018

Major: Biochemistry
Genetics

TABLE OF CONTENTS

ABSTRACT

Structure-Based Drug Discovery Against the Caseinolytic Protease from *Mycobacterium tuberculosis*

Emily Grace Tranchina
Department of Biochemistry and Biophysics
Texas A&M University

Research Advisor: Dr. James C. Sacchettini
Department of Biochemistry and Biophysics
Texas A&M University

Antibacterial drug resistance is a major factor in the increasing rate of failure to successfully treat diseases such as tuberculosis (TB). The World Health Organization estimates that in 2015 a total of 10.4 million new cases of TB were diagnosed. Of which, 480,000 cases were resistant to standard first-line therapies - isoniazid and rifampicin – called multidrug-resistant TB (MDR-TB). A subtype of MDR-TB has resistance to more aggressive chemotherapy treatments that comprise the second line of TB drugs. This subtype is called extensively drug-resistant TB and constitutes 9.5% of MDR-TB diagnoses.

Genomic studies of *Mycobacterium tuberculosis* (*Mtb)* - the causative agent of TB, have revealed a multitude of new targets for antibiotic therapies. One such target is the caseinolytic protease (ClpP1P2) and associated ATPase adaptors (ClpC1/X). The caseinolytic protease complex, ClpC1P1P2 is essential for maintaining proteostasis in *Mtb* via the degradation of misfolded and SsrA-tagged proteins identified by the ClpC1/X ATPase adaptor. Using biochemical methods and high throughput screening (HTS) technology we have characterized the inhibitory characteristics of thousands of compounds against the components of the caseinolytic protein complex.

To complement this inhibition information, we grew crystals of the active ClpP1P2 protease. Future steps of this rational drug design project will include optimization and refinement of validation assays, as well as inhibitor soaking of ClpP1P2 crystals to resolve the binding of inhibitors on a molecular level. This information will be used to make rational refinements to the structure of inhibitors to increase their specificity, inhibitory characteristics and safety as therapies for the treatment of TB.

DEDICATION

I would like to dedicate this thesis to the countless individuals who work in the shadows, who make small strides that cumulatively effect positive change in the world. Thank you.

ACKNOWLEDGEMENTS

I would like to thank my mentors Dr. James Sacchettini and Dr. Anup Aggarwal for all of their assistance, mentorship and advice over the course of this research project. Special thanks to Thomas Snavely for his work and help in creating and culturing the CRISPRi strains. Additional thanks to all past and present members of Sacchettini lab for offering their support and sharing their lab with me.

Many thanks to all of my professors at A&M who have nurtured my love of learning and helped introduce me to a wonderfully complex world of new ideas. Special thanks are given to the Biochemistry and Biophysics department for their support and encouragement of my research activities over the course of my college career.

An especially large thank you is due to all of the friends that I have made in my college career. Your help, support and humor have been one of the best parts of my college experience. Thank you for pulling me out of the weeds when necessary.

Finally, an enormous thank you to my family for their endless support and encouragement to pursue my interests.

CHAPTER I

INTRODUCTION

Tuberculosis (TB) is a disease caused by a member of the Actinobacteria genus called *Mycobacterium tuberculosis (Mtb)* (1). TB, sometimes called consumption, is characterized by degradation of lung tissue as it is colonized by *Mtb*. Common symptoms include chest pain and the presence of blood in sputum. However, *Mtb* can spread from the lungs and invade multiple organ systems and wreak further havoc on the body. The disease is spread primarily through airborne methods, infecting close family members, especially immunocompromised individuals (2).

TB is a Global Epidemic

In 2016, the World Health Organization released a report classifying TB as a global epidemic. Approximately 10.4 million new cases of TB develop each year. Concurrently, 480,000 of the 10.4 million individuals are infected with *Mtb* strains resistant to standard TB therapies isoniazid and rifampicin. These individuals are said to be infected with multidrug resistant TB (MDR-TB). A sub-group of MDR-TB, totaling 45,600 individuals have wide resistance to first and second line TB therapies. These individuals are said to be infected with extensively drug resistant TB (XDR-TB) (2). The standard TB therapy consists of at least two treatment stages and lasts for a minimum of 10 months (3). The current success rates for non-MDR-TB are greater than 85%. However, the upper bound of success for the same therapy in cases of MDR-TB and XDR-TB is 50% and 40% respectively (3).

Drug Discovery and Identification of New Drug Targets

Explanations for why the curative rates of TB are lower than desired cite a wide range of factors. Primarily sub-optimal patient compliance, cost factors, side effects and drug resistance are identified (3,4). Currently, a large drug discovery effort is being made to identify novel antibacterial agents that can be developed for use as standalone or synergistic TB therapies (3). This push is desperately needed as there are currently only nine TB drugs with favorable metrics in late stage clinical trials (3).

While developing drug therapies in general is a challenge, developing TB drugs is especially challenging. The relatively impermeable lipid envelope surrounding *Mtb* presents a large hurdle in drug delivery and consequently slows drug development efforts. As a result, lead compounds are generally tested for whole cell activity before structural refinement occurs (3-5). For a drug to be effective it must prevent an essential biological process. For example, the TB therapeutic rifampicin inhibits RNA transcription and ultimately protein translation, by binding to a subunit of the *Mtb* RNA polymerase (4). Other targets include essential biosynthetic pathways for fatty acids, cell walls, proteins and energy (3,4). Recently, genomic techniques have been used to identify additional gene and protein targets. Notably the genes encoding the caseinolytic protease– *clpp1*, *clpp2* and adaptor ATPases associated with diverse cellular functions (AAA^+) genes – *clpc1 and clpx* were all identified as essential (6). These proteins were also found to be essential (7-9).

Functions of the Caseinolytic Protease

The caseinolytic protease (ClpP) is found in both prokaryotes and eukaryotes (10). It serves many functions but generally is involved in maintaining cellular proteostasis by selective degradation of SsrA-tagged or disordered protein substrates (8,10).

In *Mtb*, ClpP proteins and AAA^+ adaptors are essential proteins that play a key role in proteostasis, stress adaptation and virulence (7,8,10-12). Members of the ClpP family are defined by a barrel-like tetradecamer composed of two stacked seven-membered rings (10).

The ClpP AAA^+ adaptors generally form hexamers and are characterized by an ATPase domain connected to a smaller unfoldase domain (12). The ClpP tetradecamer and the hexameric AAA^+ adaptor associate axially to form the Clp complex (7). Briefly, proteolysis occurs via a two-step process. First, the AAA^+ adaptor utilizes the energy yielded by ATP hydrolysis to unfold protein substrates recognized by its unfoldase domain (12). The unfolded protein is then threaded into the core of the ClpP tetradecamer where is it is hydrolyzed by any one of fourteen-serine protease active sites located in the proteolytic barrel (10).

The Caseinolytic Protease in TB and Current Drug Discovery Projects

ClpP in *Mtb* is composed of two polypeptide chains – ClpP1 and ClpP2. Crystallographic studies have revealed that two forms of the ClpP proteins can be observed, a homotetradecamer (ClpP1P1 or ClpP2P2) and a heterotetradecamer (ClpP1P2) composed of two homoheptameric rings (13,14). Computer modeling, later supported by *in vitro* testing, indicated that the catalytically active ClpP protease in *Mtb* is the ClpP1P2 complex (13). Interestingly ClpP1P2 has not been isolated in an active conformation *in vivo* and can only be formed *in vitro* using a N-blocked hydrophobic dipeptide activator (13,15). Independent ClpP1P2 complexes have peptidase and protease activities, but when co-incubated with ClpC1 *in vitro* they exhibit enhanced proteolytic activities. It is not definitively known how ClpC1 interacts with the caseinolytic complex to enhance proteolytic activity. A theory based on *in silico* modeling has been proposed that the AAA^+ adaptors ClpC1 and ClpX preferentially associate with the ClpP2 face of the ClpP1P2 barrel to form the caseinolytic complex - ClpX/C1P1P2 (16,17).

The caseinolytic complex in *Mtb* represents an intriguing drug target because alteration in the function of any component of the complex results in the lethal perturbation of proteostasis (5). It has been well-demonstrated that compounds that act as inhibitors or activators of the complex have bactericidal effects (5,8,15,17-19). Few inhibitors of the ClpP1P2 protease have been identified (18,19). Most identified lead compounds stimulate ClpP1P2 proteolysis by mimicking the interactions of ClpC1 with the caseinolytic protease, promoting dilation of the proteolytic barrel and resulting in an increased rate of non-specific proteolysis (5,17,20-23). Many compounds with activity against the caseinolytic protease are derived from natural products. Our collaborators identified ilamycin, as an unrefined but promising lead antibiotic with selective bactericidal properties against *Mtb* at low doses (24-26). They identified that ilamycin is active against the caseinolytic complex and synthesized five ilamycin analogs.

Our lab sought to characterize the bactericidal activity of these five ilamycin analogs as well as other compounds against the caseinolytic complex and its functional components using drug discovery techniques. We have refined primary high throughput screening (HTS) assays to characterize changes in the activities of the caseinolytic complex and its functional components in the presence of potential inhibitors. These primary HTS assays and refined validation assays can also be used to identify which activity of the caseinolytic complex these compounds target. We also have optimized crystal growth conditions and solved a crystal structure of ClpP1P2 using crystallographic techniques.

Our efforts have used each of these drug discovery approaches to identify and understand the molecular basis of the most potent inhibitory compounds against the caseinolytic complex. These insights will ultimately be used to optimize the potency of these inhibitors so that they may become new TB therapies.

CHAPTER II

METHODS

Protein Expression and Harvesting

ClpP1 and ClpP2

E.coli BL21-Gold *clpp* knockout strains were used to express mature versions of *Mtb* ClpP1 and ClpP2 proteins with C-terminal 6xHis- and Myc-tags. These mature proteins represent the products of autocatalytic N-terminal processing of 6 and 12 residues respectively as described previously (13). Briefly, Difco LB Broth (Miller) was prepared and sterilized. Carbenicillin (Caisson Labs, #C033-100GM) was used at 100mg/L to select for transformed *E.coli*. Starter cultures were prepared and grown overnight for at least 12 hours. Culture flasks were spiked with carbenicillin at 100mg/L and were inoculated with the described starter culture at 1ml culture/L. The transformed *E.coli* were cultured in a 37°C shaker set at 220 rpm. After 4 hours, the temperature was dropped to 30°C and induction was started with 0.5mM Isopropyl β-D-1-thiogalactopyranoside (Gold Biotechnology, #I2481C100). Protein expression occurred for 16-18 hours.

ClpC1

Expression and harvesting of ClpC1 was performed as described above with the following modifications. The same strain of *E.coli* mutants were used to overexpress the full length ClpC1 protein. Kanamycin at 50mg/mL (Gold Biotechnology, #K-120-100) was used as the selective marker. The temperature was changed to 16°C for induction.

Protein Purification

All procedures were conducted at 4°C, or when applicable- on ice.

ClpP1 and ClpP2

Three buffer mixtures were used in the purification of ClpP1 and ClpP2 proteins (13). Buffer A was composed of 100mM KCl, 50mM potassium phosphate buffer – pH 7.6, 1mM $MgCl_2$ and 10% glycerol. Buffer B was buffer A with 250mM imidazole. The gel filtration buffer was 100mM KCl, 25mM potassium phosphate buffer – pH 7.5, 1mM $MgCl_2$ and 5% glycerol. Sartorius Vivaspin® Turbo 15 30K molecular weight cutoff (MWCO) concentrators were used to concentrate ClpP1 and ClpP2 proteins by centrifugation at 4,400 rpm.

Cells pellets harvested via centrifugation at 5,000 rpm (6,238x*g*), were lysed in a chilled lysis buffer solution (buffer A + 10%B) using a Microfluidizer LM20 (Microfluidics) at 20,000 psi. The lysed cells were then centrifuged at 17,000 rpm (39,800x*g*) for 45 minutes. The supernatant was filtered through a 0.45μm membrane and purified using affinity chromatography. The supernatant was run through two (ClpP2) or three (ClpP1) connected 5mL pre-packed Nickel columns (HisTrap™ FF, GE Healthcare Life Sciences) previously equilibrated with lysis buffer using a peristaltic pump. The overexpressed protein was eluted into 5mL fractions from the nickel column using a stepwise gradient of buffer B (10%, 20%, 50%, 100%) and buffer A using an ÄKTAprime protein purification system.

Using UV-absorbance chromatographs generated by sensors in the ÄKTAprime, representative fractions were selected and analyzed using 4-20% SDS-PAGE. Briefly, samples were mixed with loading dye, boiled for 5 minutes, centrifuged at 14,000x*g* for 30 seconds and allowed to cool for 10 minutes before loading. Gels were run at 180V and then stained with Coomassie blue overnight. Comparison to the molecular weight marker allowed identification of which samples to pool and concentrate for size exclusion purification on a HiPrep 26/60

Sephacryl S-200 HR (GE Healthcare Life Sciences) column and ÄKTAprime system. ClpP1 eluted as a monomer while ClpP2 eluted as a tetradecamer.

Resulting fractions were analyzed with 4-20% SDS-PAGE gels and concentrated. Protein concentration was monitored with a Nanodrop 1000 spectrophotometer at 280nm. All proteins were concentrated to >7mg/mL, flash frozen in liquid nitrogen, and stored at -80°C.

ClpC1

Purification of ClpC1 was performed as described above with the following modifications. Three buffers were used in the purification of ClpC1 protein. Buffer A was composed of 50mM Tris – pH 7.8, 100mM KCl, 4mM $MgCl_2$ and 10% glycerol. Buffer B was buffer A with 250mM imidazole. The gel filtration buffer was buffer A without $MgCl_2$. Corning® Spin-X® UF 100K MWCO concentrators were used to concentrate the ClpC1 protein. ClpC1 eluted as a monomer after size exclusion chromatography.

ClpP1P2

The ClpP1P2 complex was formed in the presence of 5mM N-blocked dipeptide activator zLL (Bachem, CAS 7801-71-0) as described previously (15). Briefly, approximately equimolar aliquots of purified ClpP1 and ClpP2 were incubated at room temperature in 30mL of gel filtration buffer supplemented with 5mM zLL for 3 hours. The complex was concentrated to ~12mg/mL using a Vivaspin 10K MWCO concentrator. Subsequently the complex was purified with a HiPrep 26/60 Sephacryl S-200 HR (GE Healthcare Life Sciences) column equilibrated with gel filtration buffer supplemented with 5mM zLL. The appropriate fractions were concentrated to ~12 mg/mL using a 100K MWCO Corning® Spin-X® UF concentrator. Then the ClpP1P2 complex was flash frozen and stored at -80°C.

Enzymatic Assays

ATPase Assay

The ATPase assay was performed as described previously (13) with the following modifications using a Varioskan™ LUX multimode microplate reader (Thermo Scientific™). Reactions were conducted in a 96-well format at 30°C. Reaction progress was followed by monitoring absorbance of nicotinamide adenine dinucleotide (NADH) at 340nm. The optimized reaction volume of 100µL was composed of 200nM ClpC1, 1% DMSO, 400µM adenosine triphosphate (ATP), 10 units of pyruvate kinase (PK) and lactate dehydrogenase (LDH) enzymes, 200µM NADH, 500µM phosphoenolpyruvate (PEP) and 5mM $MgCl_2$ in buffer A. Plates were prepared by use of a CyBio® Well vario (Analytik Jena AG). Each reaction was started by adding ATP to the reaction mixture and monitored for a half-hour.

Peptidase Assay

The peptidase assay was performed as described previously (5,15) with the following modifications using a POLARstar® Omega plate reader (BMG LabTech). Reactions were conducted in 96 and 384-well formats at 26°C. Reaction progress was followed by monitoring the fluorescence of 7-methyl-4-aminomethyl coumarin (AMC) at 460nm (excitation at 380nm). The optimized reaction volume of 100µL was composed of 100nM ClpP1P2, 5% DMSO, 10µM of the fluorogenic substrate Z-GGL-AMC (Bachem, CAS 97792-39-7) and buffer A supplemented with 5mM zLL. Reactions were run for three hours and were started by adding ClpP1P2 protein.

Protease Assay

The protease assay was performed as described previously (5,15) with the following modifications using a POLARstar® Omega plate reader. Reactions were conducted in 96 and

384-well formats at 26°C. Reaction progress was followed by monitoring the fluorescence of fluorescein-casein (FITC-Casein) at 525nm (excitation at 490nm). The optimized reaction volume of 100µL was composed of 100nM ClpP1P2, <5% DMSO, 10µM of the fluorogenic substrate FITC-Casein (Thermo Scientific™, Cat. 23267) and gel filtration buffer supplemented with 5mM zLL.

Each reaction was started by adding ClpP1P2 protein and monitored for two hours. The same assay was performed with the ClpC1P1P2 protein which was assembled in the following manner. Purified ClpC1 at 500nM and ClpP1P2 were allowed to assemble as ClpC1P1P2 in the 96-well plate containing 5mM zLL gel filtration buffer supplemented with 2mM ATP and 5mM $MgCl_2$ for 25 minutes. The reaction was started by the addition of 10µM FITC-casein.

HTS was performed using black, flat-bottomed 384-well plates with a total reaction volume of 50µL at the optimized ClpP1P2 protease HTS conditions. Briefly a CyBio® Well vario was used to aliquot ClpP1P2 protein, substrate and buffer and compound into each well. Plates were read at 0 and 2 hour time points using the POLARStar® Omega plate reader. The raw fluorescence data was processed with the use of the Collaborative Drug Discovery software.

Percent inhibition was determined by first calculating the change in fluorescence in each well after 2 hours. Then the following formula was used to calculate percent inhibition, 100*(compound – DMSO)/(positive control – DMSO). When applicable, percent activity was calculated using the following, 100*[1 - (compound – DMSO)/(positive control – DMSO)].

Protein Crystallization

Two-well crystal plates were used to investigate the optimal crystallization conditions of ClpP1P2. A four by four pH and precipitant grid using 0.1M Bis-Tris (6.0, 6.2, 6.5, 6.8) and 20-27.5% PEG3350 was set up using the sitting drop method (15). The 2µL drops consisted of equal

parts protein and well-solution. Crystals were grown on a vibration free table at 17°C for one month. At which point the plates were moved to standard shelving. The ClpP1P2 protein used for crystal screens had a purity greater than 95% and a concentration of ~6.1mg/mL.

Additional crystal growth optimization was performed with the pH and precipitant grid described above. Specific protein to well-solution conditions were set up using a mosquito LCP crystallization robot (TTP Labtech).

X-ray Crystallography

ClpP1P2 crystals were harvested, dipped in cryoprotectant composed of 20% glycerol and well-solution, and frozen in liquid nitrogen. The crystals were exposed to X-rays at Advanced Photon Source (APS) of the Argonne National Laboratory, Chicago, IL, USA. Diffraction data were collected by shooting crystals placed 300mm from the detector with 1Å X-ray beams for 2 second exposures and 1° oscillations. Protein structures were generated as previously described (17).

Growth and RNA Analysis of CRISPRi Strains

Culture of CRISPRi Strains

Three tunable CRISPRi repression strains of the components of the caseinolytic complex were generated as previously described (27). Starter cultures were grown from 20% glycerol stocks in complete Middlebrook 7H9 broth (BD Difco™, DF0713179). Complete media was composed of 7H9 broth with 10% OADC, 0.05% Tween® 80, 25μg/mL of Kanamycin and D-Pantothenic acid. Oleic Albumin Dextrose Catalase (OADC), D-Pantothenic acid hemicalcium salt, Tween® 80, and anhydrotetracycline hydrochloride (ATc) were all acquired from Sigma-Aldrich (M0678; P5155; P4780; 37919). Strains were grown with constant shaking in a normal aerobic atmosphere at 37°C.

Absorbance readings of CRISPRi cultures using 1 mL cuvette were obtained at 600nm (OD_{600}) at 24-hour intervals to measure growth as a function of turbidity. Growth curves were also obtained using a black Greiner 96-well flat-bottom microplate containing 200µL per well.

Induction and Growth of CRISPRi Strains

After uninduced parent cultures for each strain were grown to an OD_{600} of 0.8-1, two daughter cultures were started with aliquots of the parent culture. In one of the daughter cultures, the repression of *clp* genes was induced by the addition of ATc at a final concentration of 100ng/mL. Induction strains were grown as described above.

Mtb CRISPRi Lysis, RNA Extraction and Analysis

Cultures at an OD_{600} of 0.8-1 grown for 24-hours in 1% glycine were harvested by centrifugation at 3,200 rpm and removal of the supernatant. Pellets were suspended in 1mL of chilled RLT buffer (Qiagen, 79216) and transferred to 2mL conical screw-top tubes with an O-Ring Cap filled halfway with 0.1mm diameter Zirconia beads (BioSpec Products, 11079101z).

The sealed tubes were placed in a bead beater (BioSpec). Cells were lysed by completing five cycles of 45 seconds of bead beating immediately followed by 5 minutes of cooling on ice. Then the sealed lysis tubes were centrifuged at 2,000 rpm for 30 seconds to pellet the beads. The supernatant was transferred to a sterile 1.7mL DNA/RNAse free microcentrifuge tube and spun for 1 minute at 14,000 rpm. The supernatant was transferred to a new microcentrifuge tube and stored on ice. RNA extraction was performed using the Qiagen RNeasy Mini Kit according to included instructions.

RNA Analysis

The quality and concentration of isolated *Mtb* RNA was analyzed with a Nanodrop 1000, a Qubit Flourometer, and an Agilent 2100 Bioanalyzer according to each manufacturer's instructions.

CHAPTER III

RESULTS

Optimization of Biochemical Assays for HTS

ClpC1

By coupling the activities of pyruvate kinase (PK) and lactate dehydrogenase (LDH) we were able to measure ATPase activity. This coupled reaction scheme is based on the ATPase activity of the protein of interest. The product of ATP hydrolysis, ADP is used as substrate in the reaction catalyzed by PK. PK uses the substrates PEP and ADP to yield pyruvate and ATP. The pyruvate product is then a substrate for LDH. LDH also requires the cofactor NADH to reduce pyruvate to lactate – oxidizing NADH to NAD^+ in the process. NADH strongly absorbs at 340nm but NAD^+ does not. Using this information, the ATPase activity of ClpC1 could be monitored by the decrease in NADH absorbance.

Before this assay could be used in HTS to identify inhibitors it had to be refined across multiple parameters. First, the assay was performed with differing protein concentrations (**Figure 1A**) to ascertain which condition resulted in a linear reaction rate with adequate signal separation from the negative control. In reactions with 200nM of ClpC1 protein, the duration of the linear portion of the reaction and clear difference from the negative control were ideal for HTS. Most notably, 200nM of ClpC1 was ideal because of the increased number of reactions that could be performed with the same protein pool.

Next, the effects of dimethyl sulfoxide (DMSO) on ATPase activity were investigated (**Figure 1B-D**) because it is common for compound libraries to be suspended in DMSO. A series of reactions at different DMSO concentrations were performed to understand what suppressive

effect DMSO might have on the results of this assay. The purpose was to identify what DMSO condition resulted in minimal signal depression. As expected, the 5% DMSO treatment did affect observed reaction rates. However, the 1% DMSO condition had minimal differences when compared to a reaction mixture containing 0% DMSO. In fact, by reducing DMSO concentration to 1%, the apparent rate of ATP hydrolysis suppression was shifted from 53.2% to 1.6%.

This optimization was key to increasing both the sensitivity and specificity of this assay when screening for inhibitors. These experiments also revealed that a range of DMSO concentrations could be used and still produce meaningful inhibition data. This insight allowed for a richer diversity of compounds with variable solubility characteristics to be screened for inhibitory activity against the ClpC1 protein.

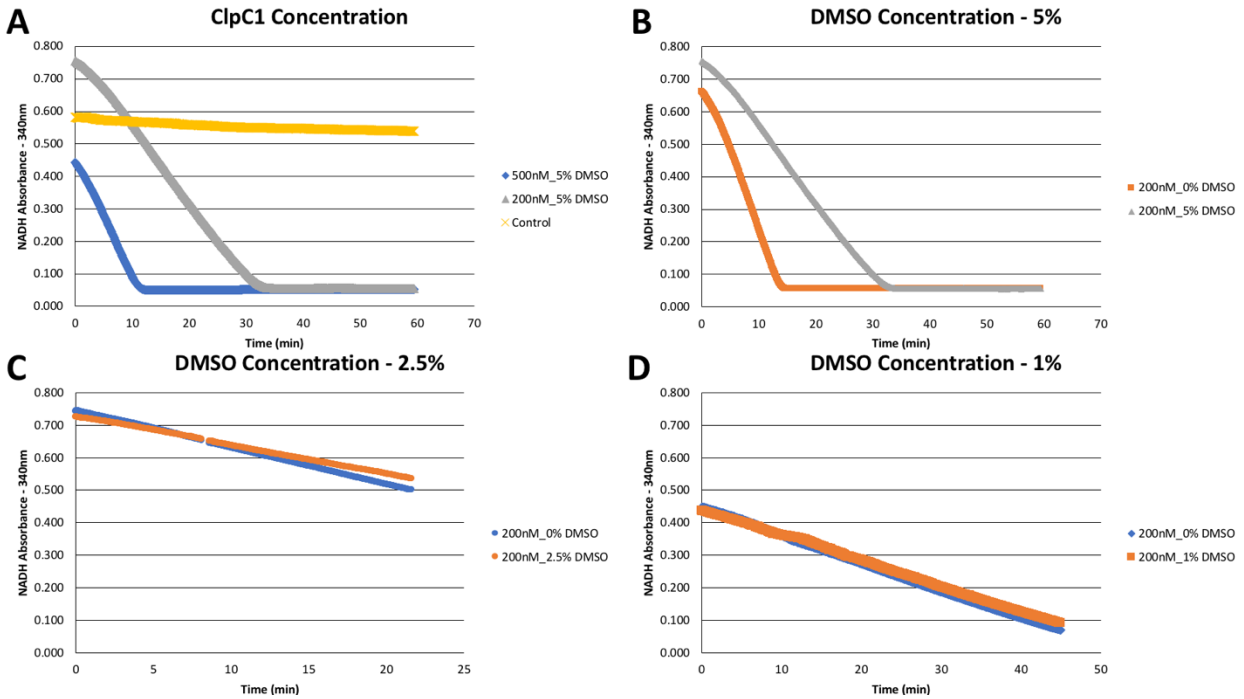

Figure 1. *ClpC1 ATPase assay optimization for HTS.* **(A)** Protein concentration. Tested signal separation and reaction duration with 500nM and 200nM of ClpC1 protein in 5% DMSO as compared to control (wells without ClpC1 protein). The assay was performed in conditions described in methods except as stated. Reactions were conducted in >1mM ATP. Results represent the average of duplicate reactions. **(B)** 5% DMSO effects. **(C)** 2.5% DMSO effects. **(D)** 1% DMSO effects.

Another key aspect of optimizing this assay was determining the apparent Km of the

ClpC1 protein for ATP. A Km is defined as the concentration of substrate required for activity of the enzyme to equal half of its maximum velocity (V_{max}). The ATP concentration in this assay was essential for yielding meaningful information in HTS. If the concentration were too high above ClpC1's Km, active site inhibitors may remain unidentified due to ATP saturation. Conversely, if ATP levels are too low, ClpC1 will not oligomerize and all compounds will look like inhibitors because the reaction rate would not occur fast enough to reveal inhibition. In reaction conditions where the ATP concentration was equal to the Km, the reaction was sufficiently fast enough to identify both allosteric and active site inhibitors.

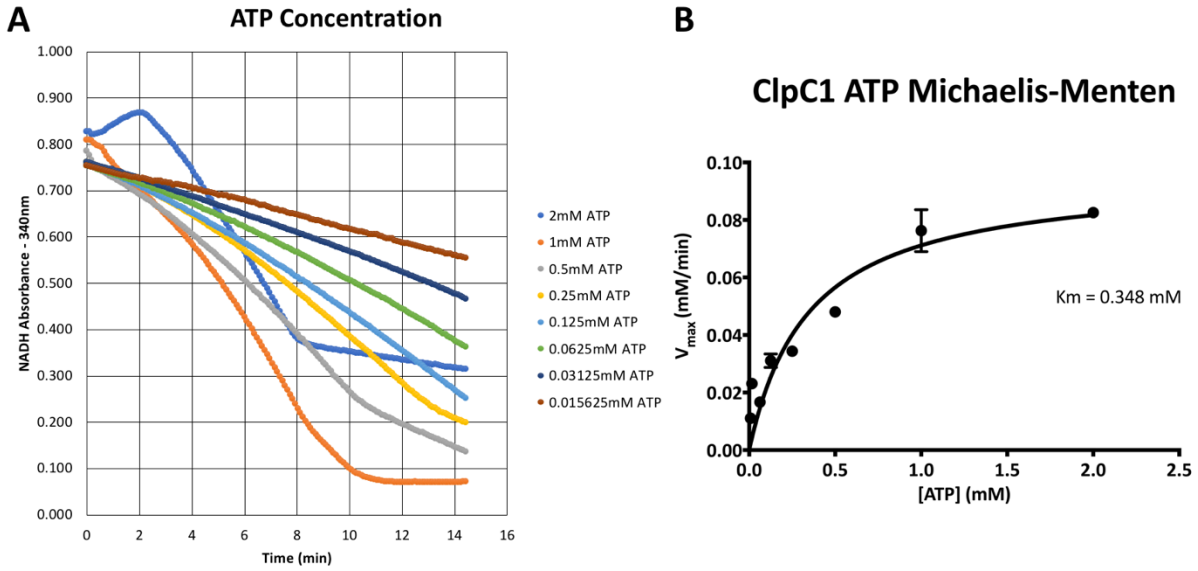

Figure 2. *Michaelis-Menten analysis of ATP concentration and ClpC1 ATPase activity*. Reactions were performed as described in methods. The reaction was conducted in the absence of DMSO with a two-fold dilution series of ATP ranging from 1mM to 16µM. Data shown are the average result of duplicate reactions. **(A)** Plot of ClpC1 ATPase activity in an ATP dilution series. **(B)** Michaelis-Menten plot of ATP dilution series data. Analysis was conducted using Prism 7's non-linear regression software. Error bars represent standard deviations.

A series of experiments (**Figure 2**) were conducted to ascertain the Km of the ClpC1 protein because it had not been characterized in the literature. We assayed ClpC1 ATP hydrolysis in a two-fold dilution series ranging from 2mM to 16µM ATP (**Figure 2A**). After Michaelis-Menten analysis, the Km was determined to be approximately 348µM (**Figure 2B**). Interestingly, this Km is nearly a thousand-fold lower than that published for the ClpC1 found in *Mycobacterium smegmatis* that shares 94% homology (16). The ATP concentration for HTS was

refined from 1mM to 400μM. This refinement also increased screening sensitivity and the number of screens possible with the same amount of reagent.

ClpP1P2

A peptidase assay is used to monitor the degradation of a peptide substrate by a protease which has both proteolytic and peptidase activities. The peptidase activity of the ClpP1P2 protease was indicated by fluorescence of 7-Amino-4-Methylcoumarin (AMC). This assay utilized Z-GGL-AMC as a reporter, where the AMC fluorophore only fluoresced when it had been cleaved from the peptide substrate. Increases in fluorescent signal indicated peptidase activity which released both amino acids and AMC into the reaction mixture.

Before this assay could be used in HTS to identify inhibitors it had to be refined across multiple parameters. First, the assay was performed at different temperatures –including 30°C and 26°C. Peptidase activity was non-existent at 30°C but was present at 26°C. This information was vital to developing HTS to minimize wasteful assays. Second, the optimal protein concentration for the assay was determined by screening 20, 50 and 100nM ClpP1P2 reactions at 26°C with 10μM of Z-GGL-AMC. An optimal concentration would yield distinct fluorescent signal as compared to the negative control- IlvC – a protein involved in isoleucine biosynthesis known to lack peptidase activity (**Figure 3A**).

Statistical analysis was performed to find the Z-factor for each protein concentration over the reaction period as compared to IlvC. A Z-factor is a statistic that measures the degree of difference between two sets of data. A result of 0.1 indicates little difference while value above 0.5 indicates a significant difference. This analysis revealed that in a 96-well format, distinct differences in readings occurred as early as 112 minutes with 100nM ClpP1P2. It also showed that the magnitude of the difference from the control stabilized after approximately three hours

(**Figure 3B**). This information was used to perform multiple, statistically valid HTS runs in a single workday.

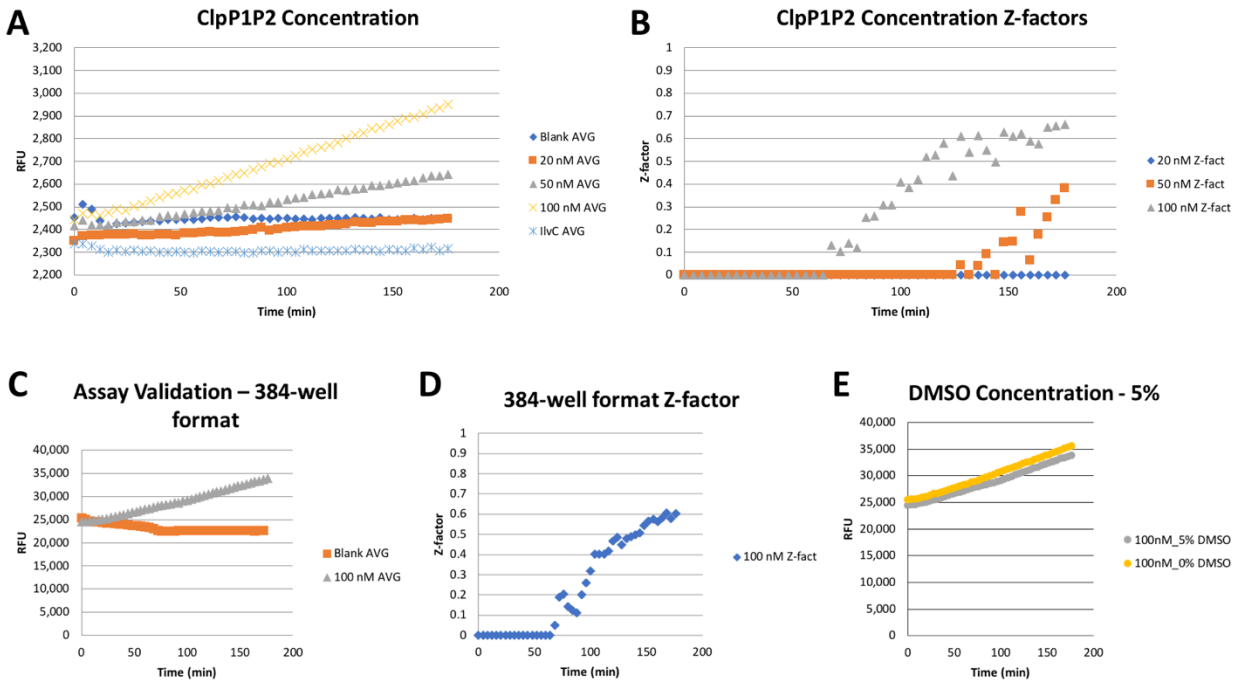

Figure 3. *ClpP1P2 peptidase assay optimization for HTS.* Reaction progress was followed by monitoring fluorescence of AMC at 460nm (excitation at 380nm). Data represent average of three reactions. Z-factors greater than or equal to 0.5 were considered significant. **(A)** Protein concentration. Tested signal separation and reaction duration with 100nM to 20nM of ClpP1P2 protein as compared to control. **(B)** Statistical analysis of ClpP1P2 protein concentration results. **(C)** Validation of the ClpP1P2 384-well format. **(D)** Statistical analysis of 384-well assay results. **(E)** 5% DMSO effects.

The optimized 96-well format reaction conditions were tested in a 384-well plate and analyzed in a similar fashion (**Figure 3C-D**). The results were similar – clear demarcation between the control was observed as early as 144 minutes. The magnitude of the difference from the control, like in the 96-well format, also stabilized after three hours (**Figure 3C**). This validation was exciting because it allowed roughly three times as many reactions to be performed in an 8-hour workday compared to the 96-well format. The effects of DMSO on the peptidase reaction were also investigated as a potential variable to optimize for HTS (**Figure 3E**). The effect of 5% DMSO on peptidase rates was minimal. No further refinement of the DMSO concentration was needed to optimize this assay for HTS.

A similar assay was optimized to monitor protease activity. All conditions – protein concentration, reaction duration, DMSO effects optimized in the peptidase assay suited the protease assay as well. The only difference between these assays was the substrate. In the protease assay the disordered protein casein was tagged with a fluorescein isothiocyanate fluorophore (FITC-casein). Reaction progress was monitored by measuring fluorescence of FITC at 525nm (excitation at 490nm). We shifted our focus to measure the activity of ClpP1P2 using FITC-casein versus Z-GGL-AMC due to the higher biological relevance of the substrate- namely its disordered nature and its diverse amino acid composition. In addition, we found that the amino acid sequence GGL is not favorably recognized and cleaved by ClpP1P2 (18).

Once we identified the optimal substrate to monitor ClpP1P2 activity, we looked to identify reference compounds that could be used as controls for inhibition of ClpP1P2 protease activity. We investigated the general serine protease inhibitor 3,4-dichloroisocoumarin (3,4-DC) as a positive control. We measured the activity of FITC-casein degradation by ClpC1P1P2 and ClpP1P2 in the presence of 200μM 3,4-DC (**Figure 4A-B)**. The extent of inhibition observed was minimal for both ClpC1P1P2 and ClpP1P2 which showed a maximum inhibition of 25% and 34% relative to DMSO respectively.

Note that in **Figure 4B** two populations of ClpP1P2 were used. “Old P1P2” was ClpP1P2 protein from a previous purification that was immediately frozen and stored at -80°C. It was thawed the day of the experiment. In contrast, “New P1P2” was purified ClpP1P2 that was left on ice for a week before use. This revealed that ClpP1P2 activity deteriorates rapidly over the course of a week. As a result, all purified protein was immediately stored in small aliquots at -80°C.

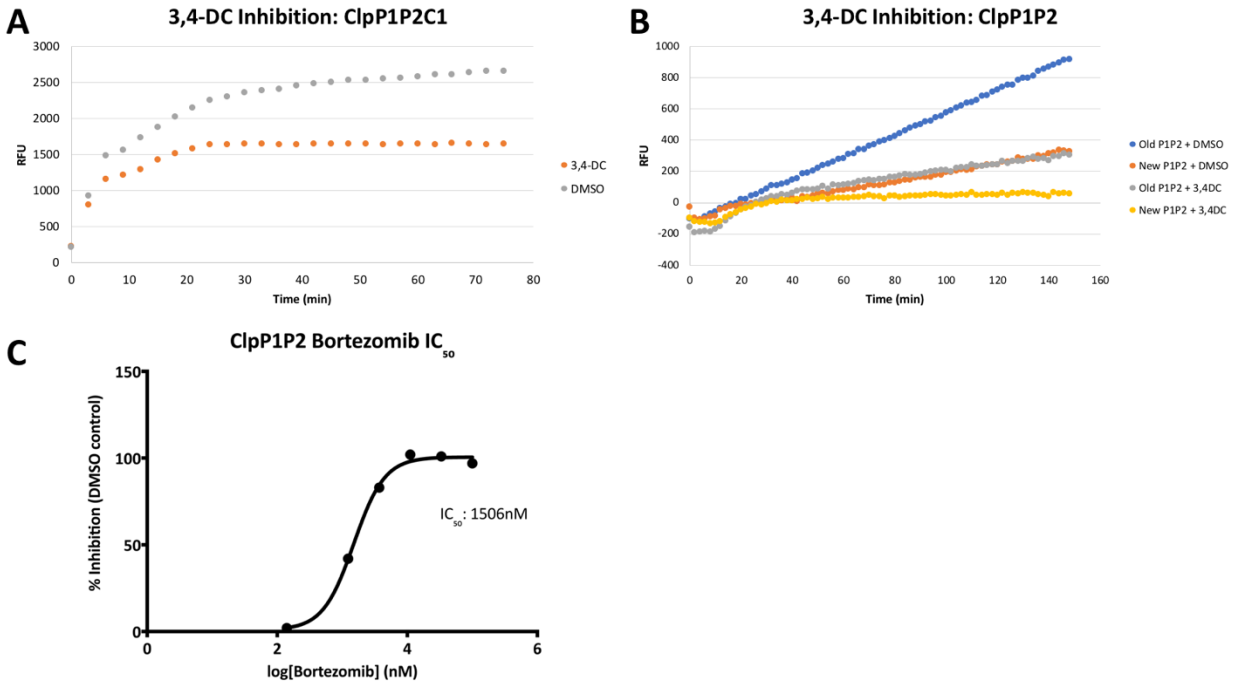

Figure 4. *Identification and characterization of inhibitor controls for ClpC1P1P2 and ClpP1P2.* The activity of the full caseinolytic complex, ClpC1P1P2 and the proteolytic barrel, ClpP1P2 was measured by monitoring fluorescence of the substrate, FITC-casein at 518 nm (excitation at 492nm). A concentration of 10µM FITC-casein was used for all reactions. **(A)** Plot of the activity of ClpC1P1P2 with 200µM 3,4-DC. **(B)** Plot of the activity of ClpP1P2 with 200µM 3,4-DC. Purified protein from two different protein pools was used to investigate the activity of 3,4-DC against ClpP1P2. **(C)** IC_{50} curve of ClpP1P2 activity with a two-fold dilution series of Bortezomib ranging from 100µM to 140nM.

We searched the literature for other potential positive controls and discovered that the peptide boronate called Bortezomib had been characterized as an inhibitor of ClpP1P2 protease activity (28). A dilution series of Bortezomib was used to identify the inhibitory concentration at which ClpP1P2 activity was halved as compared to a DMSO control (IC_{50}). Our calculated IC_{50} value of 1.5µM agreed with the literature value of 1.6µM (**Figure 4C**) (28). This result represented the final optimization step taken before performing HTS to characterize the change in activity of the caseinolytic complex or its functional components in response to a potential inhibitor.

The optimized enzymatic HTS reaction conditions described above were used to characterize the activity of 2,829 of compounds against the caseinolytic protease.

HTS

Ilamycin Analog HTS

The activities of five ilamycin analogs against ClpC1P1P2, ClpP1P2 and ClpC1 protein complexes were characterized by using a two-fold dilution series ranging from 20µM to 156nM (**Figure 5A-C**).

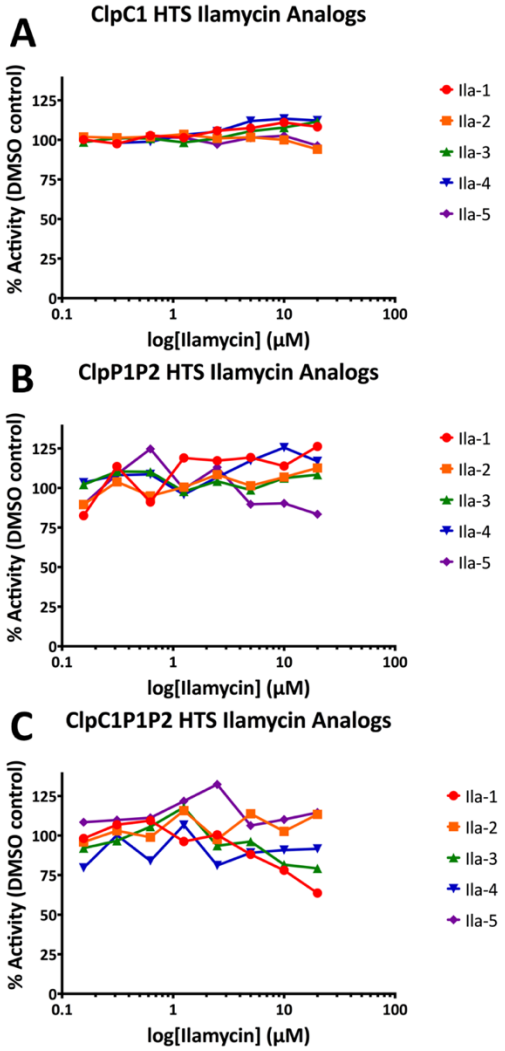

Figure 5. *Ilamycin analog HTS.* The activity of five ilamycin analogs against ClpC1, ClpP1P2 and ClpC1P1P2 was determined using a two-fold dilution series ranging from 20µM to 156nM. Wells with the protein and DMSO served the positive control. Data shown are from replicate experiments. **(A-C)** Percent activity plots of the assembled caseinolytic complex and functional components in response to treatment with five ilamycin analogs.

The percent activity analysis revealed that the five ilamycin analogs did not demonstrate significant activity against caseinolytic complex or its functional components. The lowest

observed percent activity was 67% for 20μM ilamycin-1 against ClpC1P1P2 (**Figure 5C**). Overall, we concluded that slight inhibition of ClpC1P1P2 occurred in a dose-dependent fashion for ilamycin analogs 1, 3 and 4. We did not observe inhibition of the caseinolytic complex activity when treated with ilamycins 2 and 5. These results agree with experiments performed by our collaborator (Tatos Akopian, personal communication).

As a drug discovery lab, we generally only pursue lead compounds with meaningful activity at 10 μM or below. As a result, we shifted our focus from the ilamycin analogs to search for other compounds with activity against the caseinolytic proteins.

Calibr, Sacc2 and Sanofi Compound Libraries HTS

The optimized ClpP1P2 protease activity assay was used to investigate an additional 2,829 compounds. We chose to conduct the screen using this assay due to the relative ease and speed at which a uniform, active ClpP1P2 protein pool could be purified and assembled. We also observed that compounds that exhibit activity against the ClpP1P2 also exhibited activity against ClpC1P1P2 – as seen with the modest inhibition of 3,4-DC above.

Through our participation in the TB Drug Accelerator program we are able to access compound libraries from multiple pharmaceutical companies – including Calibr and Sanofi. In addition, we have multiple home compound libraries. Our home libraries have been optimized for whole cell activity against TB, functional group diversity, and have been refined to include compounds as breakthroughs are made in TB drug discovery. The refined and diversified home library screen is called the Sacc2 Hit Plate (Sacc2).

The activity of compounds from three drug libraries Calibr (618), Sacc2 (221) and Sanofi (1900) at a concentration of 10μM against 100nM ClpP1P2 was characterized by monitoring the change in FITC-casein fluorescence at 0 and 2 hour time points. Reactions containing the

ClpP1P2 protein and DMSO were used as the positive control (blue) while wells only containing DMSO were used as the negative control (red) (**Figure 6**).

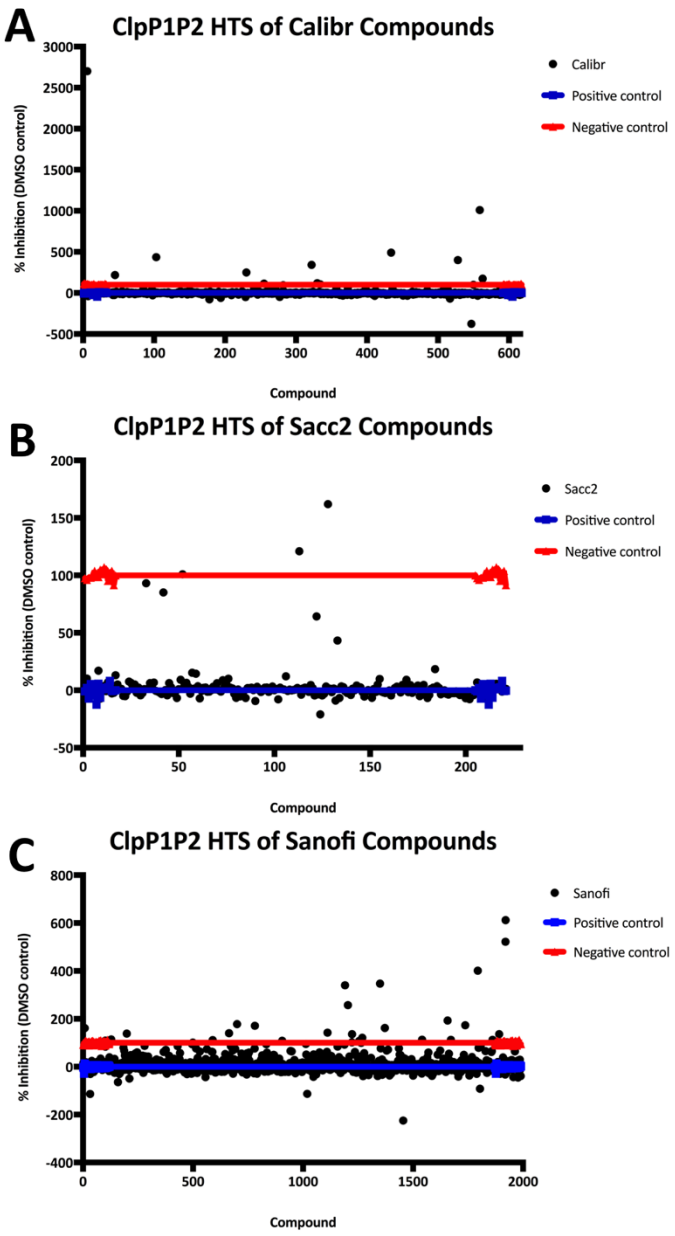

Figure 6. *Results of HTS of ClpP1P2.* The activity of compounds from three drug libraries at a concentration of 10μM against 100nM ClpP1P2 was characterized by monitoring the change in FITC-casein fluorescence at 0 and 2 hour time points. DMSO was used as the positive control while wells without protein were used as the negative control. **(A-C)** Percent inhibition plots of ClpP1P2 for each compound library.

The data were processed and plotted as a function of compound identification number and percent inhibition relative to the DMSO control (**Figure 6A-C**). Statistical analysis using Z-scores was performed relative to each compound library to identify compounds with significant

activity. These compounds are called hits or sometimes lead compounds. Hits were identified as percent inhibition values greater than 3σ from the mean for each compound library.

Our screen had 20 hits in total, resulting in an overall hit rate of 0.7%. The number of hits per compound library as listed in **Figure 6** were 4, 11 and 15 respectively. The chemical structures for the Sanofi and Calibr were not made available for our analysis. However, we were able to conduct an analysis of the hits from the Sacc2 library. We observed that the hits were chemically diverse, but other than a common feature of aromaticity and conjugation, we were unable to identify a core chemical architecture present in each hit.

Validation of HTS Hits

IC_{50} Analysis of HTS Hits

In order to better characterize the inhibitory activity of the Sacc2 hits identified in our primary screen against ClpP1P2 we performed a dose response assay (**Figure 7**).

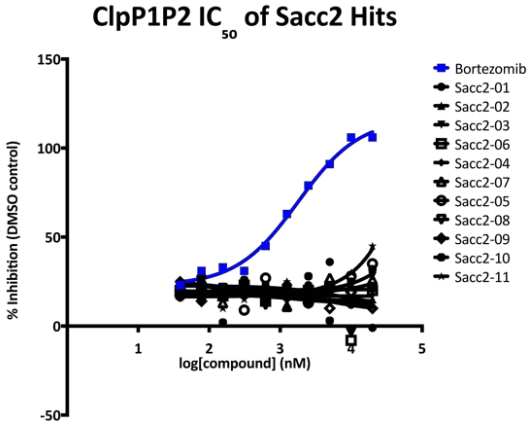

Figure 7. *Secondary screening of Sacc2 HTS hits.* IC_{50} values were determined for Sacc2 hits identified in **Figure 6B**. Bortezomib (blue) was included as a positive control for ClpP1P2 inhibition.

A two-fold dilution series ranging from 20μM to 39nM of the Sacc2 hits and Bortezomib, the inhibition control, were used in this IC_{50} assay. The resulting analysis revealed that the Sacc2 hits did not show significant inhibition of ClpP1P2 activity. This insignificant inhibition did demonstrate a concentration-dependent change in percent inhibition. The calculated IC_{50} of

Bortezomib was exactly 1.5µM as previously observed in the HTS optimization phase of this project.

Optimization of ClpP1P2 X-ray Diffraction

Two-well crystal plates were used to investigate the optimal crystallization conditions of ClpP1P2. A four by four pH and precipitant grid using 0.1M Bis-Tris (6.0, 6.2, 6.5, 6.8) and PEG3350 (20, 22.5, 25, 27.5%) was set up based on published methods (15). Long, thin rod-like crystals like those previously described, were observed in the 20% PEG3350 and pH 6.8 condition after one month (**Figure 8A**). Thin plates were also observed in the 20% PEG3350 and pH 6.5 condition after one and a half months. Crystals grew for an additional two weeks before four representative crystals – two from each condition - were harvested and sent off for data collection.

These crystals did not yield significant X-ray diffraction data. Notably, patterns indicative of salt crystals were absent from these diffraction images (**Figure 8B**). This suggests that the crystals are likely ClpP1P2 crystals but may have been too small or disordered to produce diffraction data.

Further crystal growth optimization was performed with the pH and precipitant grid described above. A total of six conditions of buffer to protein ratios (0.6/1, 0.7/1, 0.8/1, 0.8/1.2, 0.9/1.1, and 0.9/1.2) were set up using a precise robotic pipette. These crystals were grown for three and a half months.

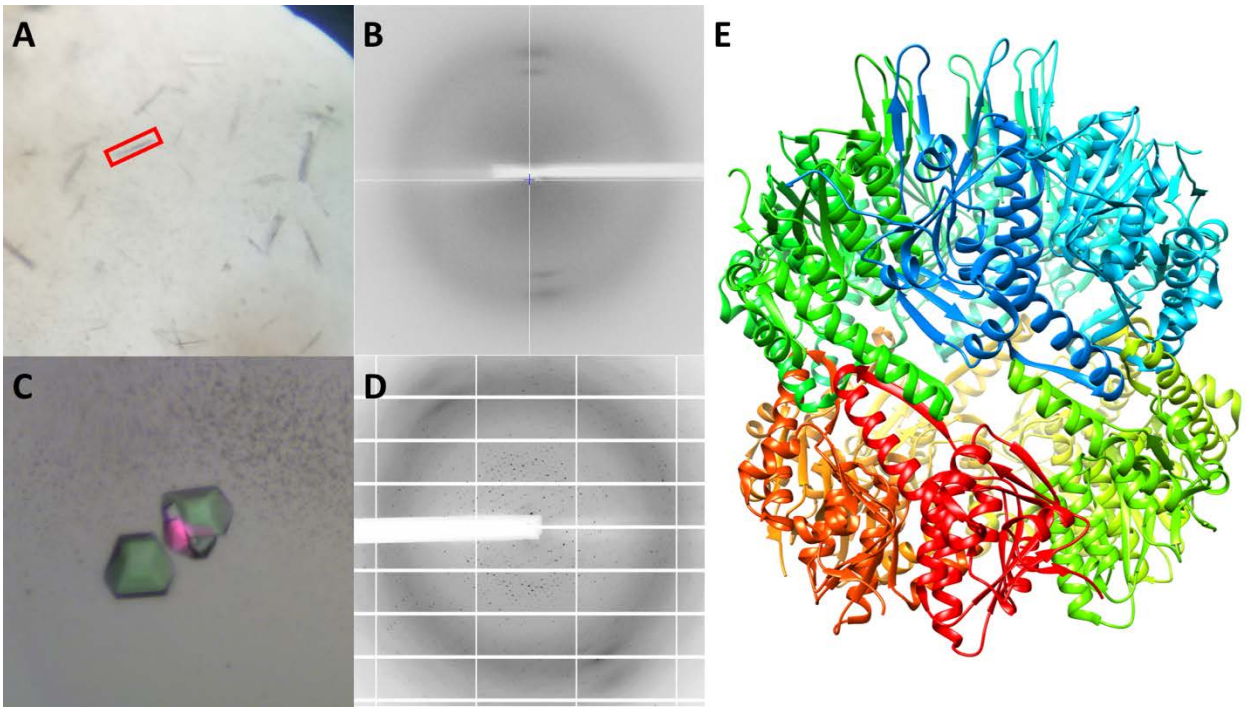

Figure 8. *Refinement of ClpP1P2 crystallization conditions and solved protein structure.* **(A)** ClpP1P2 crystals grown in 0.1M Bis-Tris, pH 6.8, 20% PEG3350. Image shows crystals under ambient light after 1.5 months growth. **(B)** Representative X-ray diffraction image of the crystal highlighted in red in **A**. **(C)** Optimized ClpP1P2 crystals. Image shows crystals under polarized light after 3.5 months growth. **(D)** Representative X-ray diffraction image of the left most crystal in **C**. **(E)** ClpP1P2 protein structure at 3.0Å resolution generated from left most crystal in **C**.

Figure 8C shows one of these optimized crystals with a protein to buffer ratio of 0.9/1.2. Interestingly these crystals exhibited a hexagonal puck-like shape. These crystals diffracted well. These diffraction data were used to produce a 3.0Å ClpP1P2 structure (**Figure 8D-E**). Our ClpP1P2 structure was identical to solved structures previously published and had 0.1Å less resolution than the highest resolution structure published in the literature (15).

Development of a CRISPRi System to Measure On-target Activity

A common component of secondary testing of lead compounds is an assay used to measure the dose-dependent growth inhibition of the target organism. Analysis of the results of this assay provide the minimum inhibitory concentration of compound needed to inhibit the growth of 50% of a population (MIC_{50}). A failure of this assay is that it measures the potency of a lead compound but does not validate or verify its on-target activity. For a vast majority of lead

compounds there is a specific essential protein target, but some compounds have secondary off-target effects.

By definition, knockouts of an essential gene cannot be generated as they would not be viable. However, with the advent of CRISPRi technology a tunable system has been developed in *Mtb* to repress, not knock out, essential protein coding genes (27). This gene repression construct is under an inducible ATc promoter. This allows the extent of the gene expression to be tuned so that the resulting protein's expression can be modulated in a range that still allows cell growth.

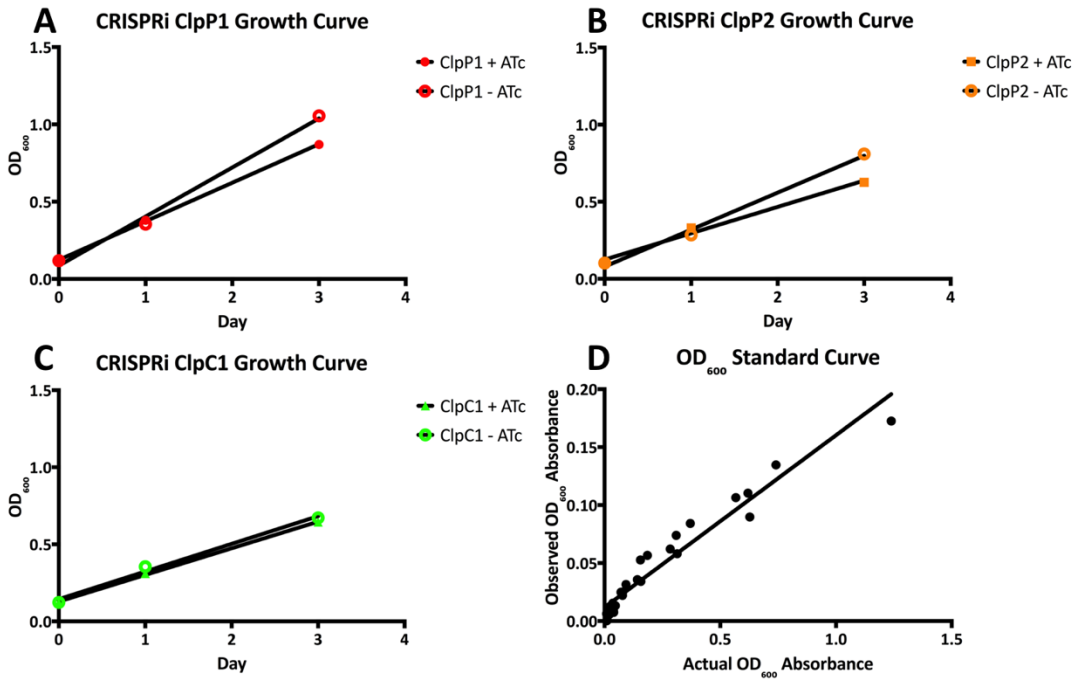

Figure 9. *CRISPRi strain growth curves and development of a high throughput OD_{600} reading technique.* Absorbance readings of CRISPRi cultures were obtained at 600nm (OD_{600}) at 24-hour intervals. **(A-C)** Growth curves of induced (+ATc) and uninduced (-ATc) CRISPRi strains taken in a 1mL cuvette. Solid black lines are the linear regressions of OD_{600} readings. **(D)** Standard curve relating OD_{600} absorbance values taken in a 96-well plate with 200µL of culture to values obtained using 1mL cuvette.

This ability to modulate gene expression provides information about which protein in the model organism is being targeted. If the compound is acting on the tunable protein target, it is expected that the MIC_{50} will decrease when the expression of the protein coding gene is decreased. If the MIC_{50} does not change or even increases, this would suggest that the compound

of interest is not acting on the repressed protein. Members in our group have used this tunable CRISPRi system on another drug-discovery project and it has shown promising results.

In order to enrich the information gained in a MIC_{50} analysis performed in traditional biochemical HTS paradigms, we worked with a collaborator to generate three tunable CRISPRi strains targeting each gene that encodes a component of ClpC1P1P2.

The uninduced CRISPRi strains exhibited the normal slow growth seen in the mc^27000 parent strain, suggesting that the genetic engineering involved in generating the strains was successful. The growth of the ClpP1 and ClpP2 CRISPRi strains under 100ng/mL ATc induction was slightly depressed relative to the uninduced strains grown from the same stock (**Figure 9A-B**). Interestingly, the CRISPRi ClpC1 strain did not show a growth defect under induction (**Figure 9C**).

The growth curves for each of the strains were generated by measuring the absorbance of 1mL of culture at 600nm (OD_{600}). Over the course of a week this process is time consuming and destroys at half of the culture sample. Instead, a high throughput way of measuring OD_{600} was developed to expedite the measurement process and minimize sample loss (**Figure 8D**). A two-fold dilution series of each CRISPRi strain and the parent mc^27000 strain was generated. To correct for differences in path length between the sample and the spectrophotometric detector, a standard curve was generated to translate OD_{600} values measured using a volume of 200µL in the 96-well format with OD_{600} values obtained using a 1mL cuvette. Dilutions of the stock culture correlated well with expected values. This allows the OD_{600} for 96 samples to be accurately determined in 1 minute while destroying 8 times as less culture.

Overall the ClpP1 and ClpP2 strains demonstrated similar growth rates. These growth rates were both faster than the growth exhibited by ClpC1 independent of induction status.

However, this information alone does not indicate gene repression. The extent of gene repression must be analyzed by RT-qPCR.

In order for the gene repression of each strain to be characterized, mRNA encoding the *clp* genes from each induction condition must be isolated. A traditional protocol using bead beating was used to lyse 10mL of CRISPRi strain culture at an OD_{600} of 1. The RNA in the lysate was purified using the Qiagen RNeasy Mini kit. The concentration and purity of this RNA was determined using spectrophotometric and fluorescence techniques (**Table 1**).

Table 1. *Determination of RNA and DNA concentration in CRISPRi RNA samples.* Purified RNA CRISPRi samples were analyzed for RNA or DNA content using three different quantitative apparatuses. The Qubit dsDNA Broad-Range (BR) dye was used to detect any dsDNA in the purified samples. RNA was purified from induced (+ATc) and uninduced (-ATc) strains.

	Nanodrop 1000 RNA (ng/µl)		**Agilent 2100 Bioanalyzer Total RNA Nano Chip (ng/µl)**		**Qubit Fluorometer dsDNA BR (ng/µl)**	
Strain	+ ATc	- ATc	+ ATc	- ATc	+ ATc	- ATc
ClpP1	378.5	1816.0	108.6608	77.77033	0.00	212.00
ClpP2	1140.9	2299.4	431.3962	1030.61	23.20	6.02
ClpC1	2035.8	2046.3	1590.981	1206.337	4.14	561.00

Initially the Nanodrop 1000 was used to determine RNA concentration. The high concentration readings were questioned because of the way this apparatus measures concentration. It determines both RNA and DNA concentration by measuring absorbance at 260nm. This lead us to question the RNA concentration and purity of each RNA sample.

A Qubit Fluorometer was used to measure the DNA content in each sample. The Qubit uses highly specific fluorescent dyes to accurately measure the RNA, DNA or protein concentration in a sample. In this case, the broad range dsDNA dye was used to detect any intact DNA in the sample. Minor amounts of DNA contamination were found in each sample.

Next, the Agilent 2100 Bioanalyzer Total RNA Nano Chip was used to determine the concentration and quality of the purified RNA (**Figure 10**). The Bioanalyzer uses densitometry, gel electrophoresis and fluorescence to quantitate RNA. While the RNA concentrations were not

identical to those found by the Nanodrop 1000, they still indicated an abundance of RNA in each purified sample. This concentration of RNA was theoretically sufficient to perform multiple RT-qPCR experiments, which generally require a minimum of 2ng of pure RNA per reaction.

The electropherogram in **Figure 10A** is similar to a UV absorbance readout – peaks correspond to the migration of different size RNA fragments in the gel. In this graph, early peaks correspond to short RNA fragments, while late peaks correspond to long RNA fragments. The results of the electropherogram indicate that the purified RNA was intact and contained a variety of nucleotide lengths. The corresponding gel is shown in **Figure 10B**. Diffuse bands are indicative of RNA degradation (see P1-, lane 2) while distinct, sharp bands denote intact RNA (see P2-, lane 3).

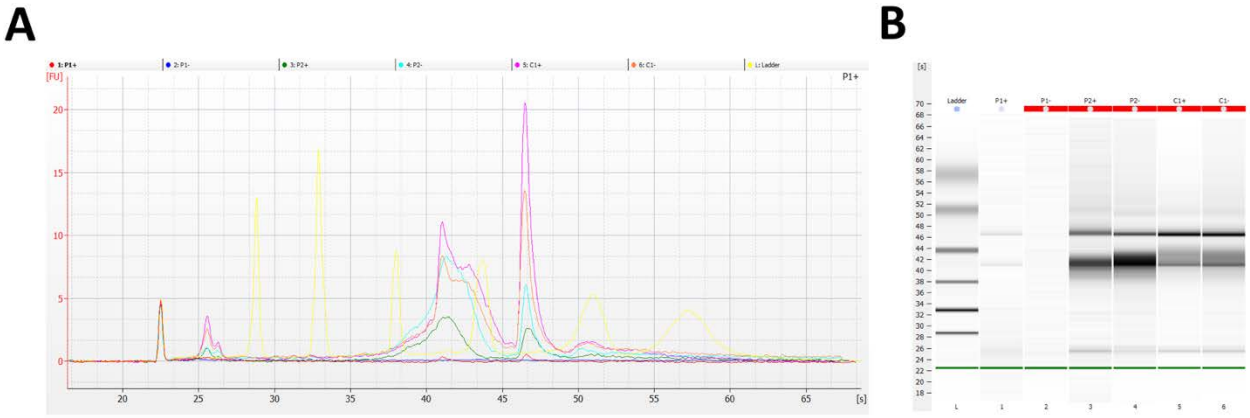

Figure 10. *Agilent 2100 Bioanalyzer Total RNA Nano Chip analysis of CRISPRi RNA samples.* Samples are labeled at the top of the gel by CRISPRi strain and induction condition (+/-). **(A)** Electropherogram graph of time (s) versus fluorescence (FU) showing the migration of RNA of varying length in the Nano Chip Gel. The extent of migration is dependent on size of the RNA fragment, larger fragments eluting last. **(B)** Gels showing the migration of purified RNA fragments in the Nano Chip Gel.

While the quality of the purified RNA using the RNA Integrity value (RIN) was unable to be determined, many sharp bands can be discerned. This suggested that RT-qPCR analysis using this RNA would be meaningful. Overall, the concentration results in **Table 1** and **Figure 10** demonstrated proof of concept and the quality of the combined bead beater lysis and RNeasy extraction approach. This suggested that RT-qPCR could be performed to validate the MIC_{50}

assay using induced and uninduced CRISPRi strains to measure the on-target activity of HTS hits against each of the caseinolytic complex proteins.

CHAPTER IV

CONCLUSIONS

Optimization of Biochemical Assays for HTS

ClpC1

The steps taken to refine the ClpC1 ATPase assay for HTS were significant. By optimizing major variables – protein, ATP and DMSO concentrations in each reaction, the overall consistency, sensitivity and specificity of the assay was improved drastically. The reduced protein concentration was extremely beneficial because of the difficulty experienced in purifying large quantities of ClpC1 protein. It took seven days to produce pure ClpC1 protein. Similarly, the short time period required to run to reaction to completion was also beneficial as it allowed for large amounts of data to be gathered in a single 8-hour period.

Future steps include optimizing ClpC1 protein expression and purification so that more HTS assays can be performed against ClpC1P1P2. In addition, a secondary validation assay will be implemented to confirm that inhibitors are active against ClpC1 and are not interfering with PK, LDH or any other components in the reaction. The previously described (13) malachite green assay holds promise for adaptation to the HTS format. Additional steps will also be taken to validate the optimized conditions in a 384-well format.

ClpP1P2

The steps taken to refine the peptidase and protease assays for HTS were significant. By optimizing the temperature and protein concentrations in each reaction, the overall consistency of the assay was improved drastically. The reduced protein concentration was extremely beneficial because it allowed more potential inhibitors to be screened from the same ClpP1P2

protein pool. Similarly, the short time period required to run to reaction to completion was also beneficial as it allowed large amounts of data to be gathered in a single 8-hour period. The validation of the 384-well format allowed the inhibitory activity of 2,829 compounds against ClpP1P2 to be determined in only three hours.

The identification of inhibitors of ClpC1P1P2 and ClpP1P2 provided a meaningful way to validate the optimized HTS assays. While the results of the Sacc2 IC_{50} response assay were underwhelming we knew that the results we observed were significant because the internal control performed as expected.

Future directions to pursue with these assays are to screen the activity of compound libraries against the peptidase and protease activities of ClpP1P2 and ClpC1P1P2 complexes. A peptidase substrate that is favorably cleaved by ClpP1P2 with a p-nitroanilide fluorophore is being developed to measure the peptidase activity of ClpP1P2. The refined peptidase assay hits will be compared to the protease assay hits to reduce the number of false positives generated by each assay. Additionally, HTS protease and peptidase assays will be performed with controls known to inhibit and stimulate the assayed activity as a measure of consistency and reproducibility.

HTS

Ilamycin Analog HTS

While the ilamycins were not found to have any significant activity against the caseinolytic proteins, they demonstrated the power of our screening techniques. The similar results produced by our collaborator demonstrates the reproducibility of our assays. These refined assays were then used to screen an additional 2,829 compounds.

Future directions to pursue with the ilamycin analogs involve additional validation to confirm the primary HTS results. The original parent compound should be included in the validated MIC_{50} assay to characterize the on-target activity of the parent, and ilamycin analogs against the components of ClpC1P1P2.

Calibr, Sacc2 and Sanofi Compound Libraries HTS

Our optimized HTS reaction parameters have allowed us to efficiently screen thousands of additional compounds for activity against the protease activity of ClpP1P2. We were able to identify 20 lead compounds for further investigation.

Future directions include performing validation of the Sanofi and Calibr hits using biochemical and genetic techniques. If possible, the chemical structure of the hits identified should be obtained to identify common structural features that could be helpful in later stages of this drug discovery project.

Validation of HTS Hits

IC_{50} Analysis of HTS Hits

The results of the IC_{50} assay did not indicate that the Sacc2 hits identified in primary HTS were hits worth pursuing further. Again, the consistency of the Bortezomib control to reproduce a consistent IC_{50} value demonstrates the power of this assay.

Optimization of ClpP1P2 X-ray Diffraction

The strides that we were able to make with our refined crystallization conditions for ClpP1P2 are extremely exciting. Ongoing efforts are being made to optimize the reproducibility of ClpP1P2 crystallization conditions so that co-crystallization or soaking with identified hits can be performed to resolve the activity of lead compounds on a molecular level.

Additional improvements in the ClpP1P2 resolution may uncover structural features that can be used to identify novel lead compounds or direct refinement of identified hits. Finally, efforts to crystallize and produce the first structures of ClpC1 and ClpC1P1P2 in *Mtb* will be made by performing crystal screens using 1000+ crystallization conditions.

Development of a CRISPRi System to Measure On-target Activity

We have used cutting edge genetic techniques to generate three strains of *Mtb* that can be used to enrich the information provided by the standard MIC_{50} assay in the drug discovery pipeline. Ongoing efforts are being made to identify and modulate the extent of *clp* gene repression using RT-qPCR analysis of purified RNA. This information can be used to validate CRISPRi strain gene repression under ATc induction. Western blots will also be performed to confirm reduced protein expression.

The beauty of the CRISPRi system is that for the first time, the relationship between target gene repression and drug sensitivity can be clearly investigated in a fast, cost effective fashion. This approach is a meaningful way to integrate genetic techniques into secondary HTS. The results from these assays will guide further research towards meaningful hits that specifically act on the caseinolytic complex. This information will in turn accelerate the TB drug discovery process and will ultimately increase the number of TB survivors.

REFERENCES

1. Barberis, I., Bragazzi, N. L., Galluzzo, L., and Martini, M. (2017) The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J Prev Med Hyg* **58**, E9-E12

2. World Health Organization. (2016) Global tuberculosis report 2016. WHO, Geneva, Switzerland

3. AlMatar, M., AlMandeal, H., Var, I., Kayar, B., and Koksal, F. (2017) New drugs for the treatment of Mycobacterium tuberculosis infection. *Biomed Pharmacother* **91**, 546-558

4. Nusrath Unissa, A., and Hanna, L. E. (2017) Molecular mechanisms of action, resistance, detection to the first-line anti tuberculosis drugs: Rifampicin and pyrazinamide in the post whole genome sequencing era. *Tuberculosis (Edinb)* **105**, 96-107

5. Famulla, K., Sass, P., Malik, I., Akopian, T., Kandror, O., Alber, M., Hinzen, B., Ruebsamen-Schaeff, H., Kalscheuer, R., Goldberg, A. L., and Brotz-Oesterhelt, H. (2016) Acyldepsipeptide antibiotics kill mycobacteria by preventing the physiological functions of the ClpP1P2 protease. *Mol. Microbiol.* **101**, 194-209

6. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**, 77-84

7. Raju, R. M., Unnikrishnan, M., Rubin, D. H., Krishnamoorthy, V., Kandror, O., Akopian, T. N., Goldberg, A. L., and Rubin, E. J. (2012) Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLoS Pathog* **8**, e1002511

8. Ollinger, J., O'Malley, T., Kesicki, E. A., Odingo, J., and Parish, T. (2012) Validation of the essential ClpP protease in Mycobacterium tuberculosis as a novel drug target. *J. Bacteriol.* **194**, 663-668

9. Raju, R. M., Jedrychowski, M. P., Wei, J. R., Pinkham, J. T., Park, A. S., O'Brien, K., Rehren, G., Schnappinger, D., Gygi, S. P., and Rubin, E. J. (2014) Post-translational regulation via Clp protease is critical for survival of Mycobacterium tuberculosis. *PLoS Pathog* **10**, e1003994

10. Kress, W., Maglica, Z., and Weber-Ban, E. (2009) Clp chaperone-proteases: structure and function. *Res. Microbiol.* **160**, 618-628

11. Kirstein, J., Moliere, N., Dougan, D. A., and Turgay, K. (2009) Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nat. Rev. Microbiol.* **7**, 589-599

12. Saibil, H. (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14**, 630-642

13. Akopian, T., Kandror, O., Raju, R. M., Unnikrishnan, M., Rubin, E. J., and Goldberg, A. L. (2012) The active ClpP protease from M. tuberculosis is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *EMBO J.* **31**, 1529-1541

14. Ingvarsson, H., Mate, M. J., Hogbom, M., Portnoi, D., Benaroudj, N., Alzari, P. M., Ortiz-Lombardia, M., and Unge, T. (2007) Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of Mycobacterium tuberculosis ClpP1. *Acta Crystallogr D Biol Crystallogr* **63**, 249-259

15. Li, M., Kandror, O., Akopian, T., Dharkar, P., Wlodawer, A., Maurizi, M. R., and Goldberg, A. L. (2016) Structure and Functional Properties of the Active Form of the Proteolytic Complex, ClpP1P2, from Mycobacterium tuberculosis. *J. Biol. Chem.* **291**, 7465-7476

16. Schmitz, K. R., and Sauer, R. T. (2014) Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Mol. Microbiol.* **93**, 617-628

17. Schmitz, K. R., Carney, D. W., Sello, J. K., and Sauer, R. T. (2014) Crystal structure of Mycobacterium tuberculosis ClpP1P2 suggests a model for peptidase activation by AAA+ partner binding and substrate delivery. *Proc Natl Acad Sci U S A* **111**, E4587-4595

18. Akopian, T., Kandror, O., Tsu, C., Lai, J. H., Wu, W., Liu, Y., Zhao, P., Park, A., Wolf, L., Dick, L. R., Rubin, E. J., Bachovchin, W., and Goldberg, A. L. (2015) Cleavage Specificity of Mycobacterium tuberculosis ClpP1P2 Protease and Identification of Novel Peptide Substrates and Boronate Inhibitors with Anti-bacterial Activity. *J. Biol. Chem.* **290**, 11008-11020

19. Moreira, W., Santhanakrishnan, S., Ngan, G. J. Y., Low, C. B., Sangthongpitag, K., Poulsen, A., Dymock, B. W., and Dick, T. (2017) Towards Selective Mycobacterial

ClpP1P2 Inhibitors with Reduced Activity against the Human Proteasome. *Antimicrob. Agents Chemother.* **61**

20. Gao, W., Kim, J. Y., Anderson, J. R., Akopian, T., Hong, S., Jin, Y. Y., Kandror, O., Kim, J. W., Lee, I. A., Lee, S. Y., McAlpine, J. B., Mulugeta, S., Sunoqrot, S., Wang, Y., Yang, S. H., Yoon, T. M., Goldberg, A. L., Pauli, G. F., Suh, J. W., Franzblau, S. G., and Cho, S. (2015) The cyclic peptide ecumicin targeting ClpC1 is active against Mycobacterium tuberculosis in vivo. *Antimicrob. Agents Chemother.* **59**, 880-889

21. Jung, I. P., Ha, N. R., Kim, A. R., Kim, S. H., and Yoon, M. Y. (2017) Mutation analysis of the interactions between Mycobacterium tuberculosis caseinolytic protease C1 (ClpC1) and ecumicin. *Int J Biol Macromol* **101**, 348-357

22. Schmitt, E. K., Riwanto, M., Sambandamurthy, V., Roggo, S., Miault, C., Zwingelstein, C., Krastel, P., Noble, C., Beer, D., Rao, S. P., Au, M., Niyomrattanakit, P., Lim, V., Zheng, J., Jeffery, D., Pethe, K., and Camacho, L. R. (2011) The natural product cyclomarin kills Mycobacterium tuberculosis by targeting the ClpC1 subunit of the caseinolytic protease. *Angew Chem Int Ed Engl* **50**, 5889-5891

23. Vasudevan, D., Rao, S. P., and Noble, C. G. (2013) Structural basis of mycobacterial inhibition by cyclomarin A. *J. Biol. Chem.* **288**, 30883-30891

24. Takita, T., Naganawa, H., Maeda, K., and Umezawa, H. (1964) The Structures of Ilamycin and Ilamycin B2. *J Antibiot (Tokyo)* **17**, 129-131

25. Kaku, R. (1963) [Studies on the antitubercular activity of a new antibiotic, ilamycin. 2. Antitubercular activity of a water-soluble derivative of ilamycin, ilamycin S]. *J Antibiot B* **16**, 99-103

26. Kaku, R. (1963) [Studies on the antitubercular activity of a new antibiotic, ilamycin. 1. Antitubercular activity of water-insoluble ilamycin]. *J Antibiot B* **16**, 93-98

27. Rock, J. M., Hopkins, F. F., Chavez, A., Diallo, M., Chase, M. R., Gerrick, E. R., Pritchard, J. R., Church, G. M., Rubin, E. J., Sassetti, C. M., Schnappinger, D., and Fortune, S. M. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol* **2**, 16274

28. Moreira, W., Santhanakrishnan, S., Dymock, B. W., and Dick, T. (2017) Bortezomib Warhead-Switch Confers Dual Activity against Mycobacterial Caseinolytic Protease and Proteasome and Selectivity against Human Proteasome. *Front Microbiol* **8**, 746